

**Slide 1 of 22**

Title Slide: Considerations for Analysis of SASS, TFS, and PFS Data

**Slide 2 of 22**

This module describes analytic considerations that should be kept in mind when using these data. Since the data in SASS are inter-related, users may want to merge data files. Some types of data file merges, such as combining data for public and private principals, are relatively simple to perform with SASS data; others require some planning. However, there are some cautions about how to merge the files and about nonresponse across data files that will also be described in this module.

In addition, this module will explain the data file considerations including the importance of unit of analysis when merging files. Lastly, this module also presents information about items with low response rates for certain variables in the 2011-2012 SASS, as well as analytic considerations when analyzing TFS and PFS data.

**Slide 3 of 22**

In SASS, not every questionnaire administered becomes its own data file.

The Public School Questionnaire (With District Items) was given to public charter schools, traditional public schools that were single-school districts, and state-run schools. All public charter schools were given the Public School Questionnaire (With District Items), because those that are a part of a public school district usually handle certain district-like functions, such as hiring and firing, independent from the district. The school-level questionnaire items in the Public School Questionnaire (With District Items) are included along with the School Questionnaire items to form the Public School data file. The district-level items for public charter schools that are not under the jurisdiction of regular school districts are added into the District file. The Charter School Analysis data file, however, contains responses from all public charter schools. Data from public charter schools under the jurisdiction of regular school districts may duplicate yet differ from those found on the District data file.

Keep in mind that related data files at the same unit of analysis or component level, such as public and private schools, may not have exactly the same number of variables. Compare the public school and private school questionnaires to see whether there are additional variables on the private school file.

Across component levels, there is the possibility of losing some corresponding sampled units due to varying response rates between components. Each questionnaire has its own individual response rate – a school may respond but the corresponding district, teacher, or principal may not have. When a one-to-one merge is done between components that do not have exactly the same number of respondents, there will be some records that cannot be merged. More detail on how to perform merges is provided later on in this module.

**Slide 4 of 22**

As a data analyst, you have decisions to make regarding how to use multiple data files in your analysis. Each data file can be analyzed separately. For example, the average hours worked by public school teachers can be generated, and a separate analysis of the average hours worked by private school teachers can be run, allowing for comparison of the average hours worked for these two types of teachers. But if your research interest is about the percentage of all U.S. teachers by the average hours worked, then the analysis requires merging the public school teacher data file with the private school teacher data file. There are two types of data file merges: the first is called concatenating files, or appending files. The second is termed one-to-one matching. How do these types of merging files differ, and how do you decide which is the appropriate type of merge for your analysis?

Let's take the first method, concatenating files. In this type of file merge, all of the observations from the second data file are added onto the same record at the end of the first data file record. To go back to the teacher example, each variable would have observations from both public teachers and private teachers. Because files merged this way do not need to be "matched," the control number does not need to be specified. However, if a variable is only on one of the data files, then the merged data file will have missing data for that variable from the data file that did not contain the variable. This method for merging files, although easy to generate, is more difficult to work with in a large and complex set of data files such as SASS. This method works best in SASS for combining the public sector and private sector data from the same component, where the variables are identical or mostly identical. Consult your statistical software package documentation for more details about how to use concatenated data files.

**Slide 5 of 22**

The other type of data file merging is called one-to-one matching of data files. For this type of merge, each data record is matched by a unique identifying variable. For SASS, the unique identifying variable is the control number. Each data file in SASS has a control number. When there is no match on control number, then the separate data records from the second data file are added to the first data file in the order of the matching variable. The entire data file for merged public and private school teachers by the matching method would have all of the public school teacher records and all of the private school teacher records, listed in control number order. If you are merging public school and public principal data files, the portion of the control number to use for matching is the school portion. When there is a match on the control number, such as for public school and public school principals from the same school, the variables from the second data file that do not match are added onto the matched record from the first data file. Matched data records will have additional data in variables that match, so for the teacher files, that means the hours worked variable will have data from both public school teachers and private school teachers. But there are no questionnaire variables in common across components, so for the public school and public school principal data files, the additional principal variables are added at the end of the school record, when the school is the first data file.

Slide 6 of 22

In merging restricted-use data files using the school control number (CNTLNUMS), both data files being merged must be sorted by the school control number prior to performing the merge. In the Stata code shown, words in italics are meant to be replaced by file or variable names that the user specifies.

When merging any of the school, principal, teacher, or school library media center files together for a given school, the school's control number, CNTLNUMS, is used to merge data files. When using Stata, a default merge variable is created during the merging of data files. The default name of this variable is **merge**. The variable **merge** identifies the various categories of data in a one-to-one merge and can be used to specify a unit of analysis. To merge the Public School District Data File with other public sector data files, the district's control number (CNTLNUMD) should be used. This variable is included on the Public School District Data File as well as the public school, principal, teacher, and library data files. The sample code provided above is correct, except that the merging variable will be CNTLNUMD. The codes for merge results, and an example of merging the public school district data file with the public school principal data file, can be accessed by clicking on the underlined screen text, 'Merge Codes Resource Document'.

The first merge category contains observations from the **principal file only** with no public school district data added because of school district nonresponse. The second merge category contains observations from the **public school district data only**, with no principal data due to nonresponse. The third category is when the merged data file contains **matching observations from the principal and the public school district data files**.

By dropping the second merge code observations, the merged data file will contain only principals, regardless of whether their district responded. No observations will remain when a district responded without a principal.

Slide 7 of 22

In merging files within SASS using SPSS, both data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. In SPSS, value labels are attached automatically during the extraction process. For the code shown on screen, words in italics are meant to be replaced by the file or variable names that the user specifies.

When merging any of the school, principal, teacher, or school library media center files together for a given school, the school's control number, CNTLNUMS, is used to merge data files.

In the SPSS syntax shown on screen, first the two data files are specified, and each one is sorted by the school control number. The file specified as *Dataset1* is the unit of analysis, so if the analysis is combining principal data with school data, and principals are the unit of analysis, *Dataset1* is the principal file and *Dataset2* is the school file. In

## Considerations for Analysis of SASS, TFS, and PFS Data

SPSS, the command for merging is called **match files**. The **table** statement that follows a forward slash in the **match files** command specifies the second data file, and the **by** statement provides the variable on which to perform the merge.

### Slide 8 of 22

To merge the Public School District Data File with other public sector data files, the district's control number (CNTLNUMD) should be used in place of CNTLNUMS. This variable is included on the Public School District Data File as well as the public school, principal, teacher, and library data files. The sample code provided above is correct, except that the merging variable will be CNTLNUMD.

### Slide 9 of 22

In merging restricted-use data files within SASS with SAS statistical software, please note that both data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge.

Here is an example of using SAS to perform a matching data file merge. Portions of this code have been 'commented out' to provide users with an explanation of what each line of code does when run. 'Commented out' code is contained within forward slashes and astericks. It is also important to note that words shown in this example in italics are meant to be replaced by the file or variable names relevant to your analysis.

In SAS syntax, the convention is to identify the data file that is the unit of analysis with the statement "in equals a," and that statement has to appear within parentheses. It can be used in a variety of ways in one-to-one and one-to-many merges. For more information on different types of merges and using the "in equals a" convention, users should refer to the SAS statistical procedures manual.

### Slide 10 of 22

Value labels are not automatically attached to SASS variables in SAS, so getting those values pulled into your dataset requires using code that can be found in the [Value Labels Resource Document](#), which can be accessed by clicking on the corresponding underlined screen text. The SAS data files (with assigned formats) on the restricted-use data files CD have assigned value-label formats. These are provided in the SAS program files ending in "underscore FMT period SAS." These format statements must be run first in order to create a format catalog to use with the formatted SAS datasets (data files ending in \_FMT.SAS7BDAT).

To merge the Public School District Data File with other public sector data files, the district's control number (CNTLNUMD) should be used. This variable is included on the Public School District Data File as well as the public school, principal, teacher, and library data files. The sample code provided above is correct, except that the merging variable will be CNTLNUMD.

**Slide 11 of 22**

As mentioned in “SASS Sample Design, Weights and Variance,” each data file in SASS has a different final weight. If you are only using one data file, such as the public teacher data file, then TFINWGT is the weight you need. But if you have followed any of the above steps to merge together more than one component, your merged dataset will have more than one final weight. While a merged school and teacher data file might require the teacher weight for one analysis, another analysis might require the school weight. In some cases it may not be obvious which weight is the proper one to use.

**It is extremely important to identify the appropriate unit of analysis and use the correct weight, or else your analytic approach may be misguided and have misleading results.** Usually, considering the target population is the best way select which weight variable to use. In other words, considering which group the analysis is being generalized to represent will help you identify the unit of analysis for your research.

If the research question is “Do rural schools have a higher proportion of experienced teachers than those in other community types?” the unit of analysis is the school, although the data on the experience level of teachers is drawn from the teacher file. The appropriate final weight is the school weight. But if the research question is “In rural schools, are experienced teachers more likely to be dissatisfied with school safety policies?” the unit of analysis is the teacher, since teachers are the ones voicing their opinions. Note that while rural schools might seem to be the target population, it is actually the teachers in those schools whose opinions are being represented. Consequently, in this second scenario, the teacher final weight should be applied.

**Slide 12 of 22**

Some specific subcategories of respondents to 2011-12 SASS fell below the IES/NCES standard for an accepted level of unit response rates (or participation rates). NCES will not publish estimates where the overall response rates fall below 50 percent until additional bias analyses are done and more information about potential bias is known. Data users who choose to generate estimates for these states are encouraged to treat them cautiously. Additional details on public school unit response rates are available for users with restricted-use data within Table 3 of the *User’s Manual for the 2011–12 Schools and Staffing Survey Volume 1: Overview (NCES 2013-330)*, which can only be accessed from the restricted-use DVD.

Unit response rates fell below 50 percent for: **public school principals in the District of Columbia and Maryland; private Jewish school principals; public schools in the District of Columbia and Maryland; private Jewish schools.**

**Hawaii** is a unique state in that it has only one school district, and that district did not respond to the 2011–12 School District Questionnaire. As a result, district estimates cannot be obtained for the state of Hawaii. SASS district weights have been adjusted for

## Considerations for Analysis of SASS, TFS, and PFS Data

the Hawaii nonresponse, thus, district estimates are still representative at a national level.

Unit response rates fell below 50 percent for **library media centers in the District of Columbia and Maryland** and for **public school teachers in City schools and in the following states: Alaska, the District of Columbia, Florida, Hawaii, Maryland, and Rhode Island**. City schools are defined as schools located in an urbanized area with 100,000 or more residents.

Lastly, unit response rates fell below 50 percent for **private school teachers in other religious schools**,

nonsectarian schools, suburban schools, rural schools, combined schools, schools with less than 200 students, schools with 750 or more students, and in the following affiliation strata: Baptist, Jewish, other religious, nonsectarian regular program, and nonsectarian special emphasis. “Nonsectarian” private schools are those that lack a religious orientation. “Combined” schools are those with grade levels at both the elementary and the secondary level, such as K-8 or 1-12. “Other religious” private schools are those with a religious orientation but not Catholic or Jewish.

For details about nonresponse bias analyses at the unit and item level, please review Chapter 10 in the *User’s Manual for the 2011–12 Schools and Staffing Survey Volume 1: Overview (NCES 2013-330)*, which can only be accessed from the restricted-use DVD.

### Slide 13 of 22

In addition, there were a few **data anomalies in specific SASS items**:

In cases where **a district reported more than 999 short-term substitute teachers** (D0450), the value was top-coded to 999. In cases where this variable was imputed and the imputed value was greater than 999, the value was top-coded to 999.

**A small number of library media centers reported more paid professional staff who held a master’s degree (M0063) or who were state-certified as a classroom teacher (M0064) than the combined total of all paid professional staff reported** (M0051–M0052, M0055–M0056). While it is unknown why this occurred, these cases were edited so that the number of paid professional staff with a degree or state certification could not be greater than the total number of all paid professional staff.

### Slide 14 of 22

In a few cases, **the number of library media center computer workstations with access to the Internet (M0076) exceeded 99**. The answer space for this item allowed for only two digits. It is possible some respondents reported 99 here when the actual number of computer workstations with internet access was 100 or greater.

During data checks, the mean for each continuous variable was computed, inclusive of imputed cases, and compared with the mean for that variable, exclusive of imputed cases. For **public and private school teachers, the mean value of T0043 (years spent teaching in both public and private schools, concurrently) was effected by imputation**. When using this variable, users may wish to restrict analyses to nonimputed cases using the imputation flag for this variable (F\_T0043).

For **private school teachers**, the means of both T0044 (**years in public schools only**) and T0094 (**number of students in a class for self-contained and team teachers**) were effected by imputation. When using these variables, users may wish to restrict analyses to nonimputed cases using the imputation flags for these variables (F\_T0044 and F\_T0094).

### Slide 15 of 22

In 2011-12 SASS, only the Private School Teacher data file had 3 variables with final-weighted response rates that fell below NCES Statistical Standards. These items included Q39e(4) – additional state certification content areas; Q72 – compensation based on student performance; and Q75 – receive a teacher pension. Note that these items are much less applicable to private school teachers than most SASS items.

For the public and private school principal data, all survey items included in the First Look Report have item response rates of 85 percent or more. All items on the public school district questionnaire had item response rates of 85 percent or more.

Given the low overall response rates for both public and private school teachers, a decision was reached to consider all significant differences observed in the comparisons conducted for the nonresponse bias analysis. NCES Statistical Standards require a nonresponse bias analysis when the level of response falls below the threshold rate of 85 percent. The reason for this is that survey respondents who skip some items may differ in key characteristics from survey respondents who supplied data to those items. If that is true and if a sufficient percentage of respondents skipped items, then the data supplied may potentially be biased. Biased data are responses that may not be representative of all the target population. Response rates below 85 percent have to be analyzed for potential bias, but those lower response rates do not automatically mean that the data are biased. Bias occurs when the nonresponse is clustered in some categories of nonrespondents but not in others. If the nonresponse is widely scattered more or less randomly across all types of potential respondents, then the actual responses may not be biased. Nonresponse adjustments were designed to reduce or eliminate bias. Nonresponse bias analysis looks at whether the nonresponse adjustments handle the potential bias completely or not. A basic overview of the bias analysis results is provided in each of the SASS First Look Reports.

**Slide 16 of 22**

SASS is designed to support explicit sampling strata estimates that were discussed in the module titled “SASS Sample Design, Weights, Variance, and Missing Data” and in Chapter 4 of the *2011-12 SASS User’s Manual, Volume I* (NCES 2013-330), which is only available via the restricted-use CD-ROM. While SASS stratifies by state for traditional public schools, principals, and library media centers, that does not mean that subgroups below the state level are supported in the same way. For any data analysis, it is imperative to compute the correct standard errors for all estimates computed. Analysts who do not do this may commit either a Type I or Type II error, finding data significant that are not statistically different, or failing to identify results as statistically significant. The variance section of the “SASS Sample Design, Weights, Variance and Missing Data” module covers how to correctly compute standard errors in three of the most commonly used statistical software packages.

Potential survey respondents to SASS are informed that all survey results will be published only in aggregate statistics. While there are no names recorded in the SASS data files, some data elements could be used to potentially identify specific survey respondents. As a data analyst, you have an obligation not to identify specific respondents – that includes schools and districts – in any research results or papers you publish. You must formally agree to this under penalty of law when signing into PowerStats or as part of the restricted-use data licensing process.

**Slide 17 of 22**

These modules describe how to use the most recent 2011-12 SASS. Earlier versions of SASS may have sampling changes, content changes, or processing changes not covered here. Each round of SASS has its own Data File User’s Manual or Data Documentation applicable to that data collection. Crosswalks do exist for comparing item content over time but each user is well advised to study questionnaire wording directly from one data collection to another.

A small proportion of the same schools is included in SASS sample from one data collection to the next but is not sufficient for any form of longitudinal or pseudo-longitudinal analysis. Samples of principals and teachers are independent in each data collection of SASS.

Additionally, geographic identifiers below the state level have changed over time across all NCES surveys. Many of these changes are due to statistical policies directed by OMB.

**Slide 18 of 22**

Sample surveys rely on achieving an acceptable level of response. Although there is no hard and fast rule as to when that occurs, if response rates fall below 50 percent, a decision has to be made about the quality of the data. For 2012-13 Teacher Follow-up Survey (TFS) private school teachers, NCES survey managers determined that the response rate was low enough that the reported private school teacher data could not



## Considerations for Analysis of SASS, TFS, and PFS Data

be released along with the data from public school teachers in TFS. A more detailed nonresponse bias analysis will determine whether there is sufficient representativeness among the private school teacher data to be released. As mentioned previously, if the nonresponse occurs more or less at random, then the low response rates would not affect the quality of the reported data, since nonresponse adjustments would be able to correct for it. However, if the nonresponse occurs much more often among certain types of private school teachers, such as in very small schools, then it would have to be determined whether any of the private school data could be released.

The First Look report and data files released in 2014 with 2012-13 TFS contain data only for public school teachers. The First Look report can be accessed by clicking on the corresponding underlined screen text.

Variables F1140 through F1156, part of item 13 on the Current Teacher questionnaire, were dropped from the Current Teacher data file due to data quality issues. These items were included in the questionnaire to assist teachers' calculations of their weekly instructional hours. However, the reported data had too many inconsistencies between these items and F1157, the total weekly hours reported in item 13 (How many hours per week are you paid to deliver instruction to a class of students at your current school?), and F1160 and F1161 of item 14 (total weekly paid hours).

Data file comparisons between Current Teacher base salary and the comparable SASS base salary did reveal some cases where:

- Stayers were full time in SASS and TFS, but base salary decreased
- Stayers were full time in SASS and part time in TFS, but base salary increased

It is important to note that no changes were made to the records with these data anomalies; however, analysts should be aware of these anomalies and use these variables with caution.

### Slide 19 of 22

The two data files for 2012-13 TFS are based on Current Teacher and Former Teacher questionnaires. The Current Teacher data file covers both stayers and movers; the Former Teacher data file covers leavers. Given the low response rates for private school teachers in 2012-13 TFS, the two data files released do not contain records for private school teachers on either the Current Teacher data file or the Former Teacher data file. This means that there are fewer records than for all respondents to TFS, and the Current Teacher variables mentioned previously were also dropped from the data file.

The 2012-13 TFS data files contain both the TFS variables and the matching 2011-12 SASS variables on each current or former teacher data record – that is, data on the SASS respondents who were sampled and responded to TFS. All TFS variables start with the alpha prefix “F” while the SASS Teacher variables start with “T.”

**Slide 20 of 22**

Concatenation is appropriate in merging the two TFS data files together, since the two data files do not contain any matches.

The same merging instructions hold for TFS as for SASS, except that the control number to use when merging TFS to SASS is the SASS teacher control number rather than the TFS control number. The TFS control number identifies the unique record within the TFS year but does not link back to the same teacher's record in SASS. The link back to SASS is provided by the SASS teacher control number CNTLNUMT, a variable found among the derived variables. Note that all of the demographic characteristics of (age, gender, ethnicity, race, and income) as well as school background variables pertaining to teachers (such as community size, enrollment size, grade level range) are found in SASS and are not collected again in TFS.

**Slide 21 of 22**

SASS has undergone a number of major content and procedural redesigns over the course of more than 20 years. Moving forward from 2011-12, SASS is undergoing another major redesign that excludes the district and school library components and redefines the content and coverage of the teacher, principal, and school components.

This major redesign will become the National Teacher and Principal Survey (NTPS). The NTPS will maintain many of the SASS teacher, school and principal items through core content and a series of rotating modules. "Core content" refers to items that will be collected every two years during each data collection. "Rotating modules" refers to administering modules or sets of related items not in every data collection but rotated into the questionnaires on a regular schedule. Other items traditionally collected in SASS will be replaced with data from other NCES sources.

Please check the NCES website for more information about NTPS methodology, content, and data collection schedule.

**Slide 22 of 22**

This module described analytic considerations that should be kept in mind when using SASS data. Since the data in SASS are inter-related, users may want to merge data files. Some types of data file merges, such as combining data for public and private principals, are relatively simple to perform with SASS data; others require some planning. However, there are some cautions about how to merge the files and about nonresponse across data files that will also be described in this module.

This module explained the data file considerations including the importance of unit of analysis when merging files. The module also presented data on items with low response rates for certain variables in the 2011-2012 SASS, as well as analytic considerations when analyzing TFS and PFS data.

## **Considerations for Analysis of SASS, TFS, and PFS Data**

Important resources that have been provided throughout the module are summarized in this slide along with the module's objectives for your reference.

This concludes the SASS dataset training. You may now click the exit button to return to the landing page.